C-O-N-F-I-D-E-N-T-I-A-L

CODIB-D-58/1
16 June 1960

UNITED   STATES   INTELLIGENCE   BOARD

COMMITTEE ON DOCUMENTATION

Statement of the Air Force Member,
CODIB Subcommittee on Mechanical Translation,
on CODIB-D-58*

I have requested the opportunity to address the committee prior to any discussion on Item 2 of the agenda in order to comment upon the attachment to Mr. Borel's memorandum distributed just prior to this meeting. My particular concern is that the IBM proposal did not explain the Air Force program with sufficient clarity and that, as a result, a number of fundamental misunderstandings appear to have arisen which the attachment reflects. In what follows, I shall deal with these misunderstandings one by one, pointing out in each case the facts as I understand them.

I can disclose to you at this time that competent data processing engineers and linguists responsible for research and development have carefully evaluated the IBM proposal. The Air Force accepts the proposal and heartily concurs in the establishment of a proposed facility.

Paragraph 2 of the attachment states that IBM only recently revealed operational capability as an explicit objective. It should be made perfectly clear here that, although the ATF**proposal was submitted by IBM, this corporation is executing part of an Air Force program which was conceived in 1955 with the aim of operational capability at the earliest possible date. In the fall of 1957, it was publicly announced that the Air Force intended to have an operational capability in November of 1961. At no point have we deviated from this schedule.

IBM has accepted selections submitted over a wide range of subject matters and has translated them on the spot before large audiences.

I might point out that the Air Force has the only MT dictionary of a size approaching that required for translating operationally. Currently, it numbers [        ]epresenting a vocabulary[        ]We do not know of any other dictionary that contains more [        ]even after several

25X1

25X1
25X1

---

*This statement by Mr. Robert F. Samson, is a comment on the Draft Intelligence Community Position on the IBM Proposal for an Automatic Translation Facility, and is incorporated by reference in the minutes of the 13 June meeting (CODIB-M-23 para. 4).
**Automatic Translation Facility

C-O-N-F-I-D-E-N-T-I-A-L

CODIB-D-58/1
16 June 1960

25X1

years' effort.  Since each added entry in a stored-program system must be accommodated in the programming, augmenting dictionaries for such systems has been openly acknowledged to be a slow and expensive procedure.  Our dictionary

Paragraph 1 of the attachment calls attention to the requirement for operation in real time.  This requirement is met only by the Air Force system. From our own experience, we know that the programming of language rules would be too expensive to achieve in-line, real-time operation with standard arithmetic computers.

Paragraph 3 of the attachment states that the Air Force has developed a special-purpose machine.  There were, I think, attempts in the IBM proposal to show that the Air Force machine system constituted a general-purpose language processor which has the logical capability of meeting the linguistic requirements.  The attachment itself, in paragraph 6, acknowledges the potential of the equipment for handling general problems in information retrieval.

Paragraph 3 also states that, at present, only word-for-word translations are being achieved.  We have repeatedly demonstrated a phrase-by-phrase capability. The dictionary entries deal with wide linguistic structures from entire clauses through whole predicates, prepositional phrases, compound substantives, and proper names to complete verb, noun, and adjective inflectional patterns.

Paragraph 3 contains still another inaccuracy, namely that the output of the current system is solely a flexo-writer, whereas, in fact, a magnetic-tape output is available for driving any high-speed printer.

Paragraph 4 of the attachment states that the Air Force approach leaves a number of language problems unresolved.  We have attacked and partially resolved all the problems cited, and I have just mentioned some specific examples.

There is an implication in paragraph 4 that the Air Force system is logic limited.  It is not generally appreciated that the logical operations required for linguistic analysis can be done more economically by table look-up than by numerical algorithms.  For those who grasp this subtle point in computer theory, it is obvious that there is no constraint upon the system's intrinsic power. Dr. E.R. Piore, Vice-President of IBM, told the CODIB and SCOMT group during their visit that IBM's chief corporate interest in the translation system was the development of machine organizations better suited to manipulating information in lexical form.

The skepticism expressed in paragraph 4 concerning the rate of further achievement is in conflict with the Air Force's evaluation of our contractor's previous performance in this area, as the contractor has never failed to meet contract schedule.

CODIB-D-58/1
16 June 1960

Paragraph 7 of the attachment states that, after considerable hardware development, the Air Force-IBM program now requires the expenditure of vast funds for intensive research and development on a method of translating languages by machine. We feel that, as evidenced by the superiority of our present output, a great deal of method is incorporated in the design of the organization of the hardware already built. Let me also point out that our examination of the IBM proposal indicates that they are requesting only in the neighborhood of

25X1

research and development on the method per se, which seems quite reasonable for a task of this importance. This amount, in any event, does not contribute substantially to the overall systems cost, the economy of which is reflected by the extremely low per-word cost estimated from the financial figures submitted with the proposal.

As for the repetition of the claim to a wider range of logical operations for a stored-program approach, this simply cannot be substantiated in computer theory. It is noteworthy that the Harvest System, the most advanced logical system under development at IBM, is being increasingly forced to resort to table look-up methods. By the same token, the capability of the photostore to perform arithmetic in a real-time application was recently demonstrated.

Paragraph 7 points out that with recent and anticipated advances in storage techniques, storage will no longer be a problem. In its proposal, IBM states its willingness to use other techniques, if and when they are shown to be more economical. But I understand that the only new storage device seriously being considered for inclusion in pending generations of computer systems is, in fact, the photostore itself.

The statement is made in paragraph 8 to the effect that the initial goal should be a pilot operation limited both as to languages and disciplines is in essence recommending Georgetown's present approach in spite of the evident broader capability of the Air Force system. It is an illusion to suppose that the range of the meaning of words used in a given paper is restricted by the discipline in which the paper is written.

Finally, paragraph 9 is unclear and should be rephrased to make it unmistakable that the ballot refers only to action on the IBM proposal.

In conclusion, we would like to underscore the fact that the Air Force-IBM program brings to the Intelligence Community and to the Nation the benefit of $7,000,000 and five years of successful research in the area of automatic language translation.

Without a doubt, the hardware techniques we are willing to donate offer the fastest practical and economical means of achieving operational status for such a facility. The hardware I refer to represents an investment of $5,000,000 and includes a print reader and composing equipment in addition to the translator itself. We are also willing to contribute the results of all linguistic research

CODIB-D-58/1
16 June 1960

25X1

sponsored jointly by the National Science Foundation and the Air Force. The
Air Force [                                                                    ]
period. So far, we have found it perfectly feasible to exploit the results of
our linguistic research on the hardware made available by the engineering effort,
and we can foresee absolutely no difficulty in this respect in the future.

It seems to me that the attachment fails, in particular, to appreciate that
IBM intends to use the Air Force results in linguistics and to pursue its own
linguistic research, as is shown by the listing in the ATF proposal of prominent
linguists who are currently engaged in studies on structural and transformational
grammars. However, linguists alone cannot solve mechanical translation, and
experts in machine organization are essential to the development. One last
point is that IBM has also invested over $2,000,000 in automatic language trans-
lation and that it would obviously be imprudent of the United States not to take
advantage of this research.